



(12) 发明专利申请

(10) 申请公布号 CN 122388187 A

(43) 申请公布日 2026. 07. 14

(21) 申请号 202610405277.5

(22) 申请日 2026.03.30

(71) 申请人 中煤科工开采研究院有限公司

地址 101399 北京市顺义区中关村科技园
区顺义园临空二路1号

(72) 发明人 吕依濛 范明建 武童

(74) 专利代理机构 北京清亦华知识产权代理事
务所(普通合伙) 11201

专利代理师 孟洋

(51) Int. Cl.

G06F 16/36 (2019.01)

G06F 18/25 (2023.01)

G06N 5/02 (2023.01)

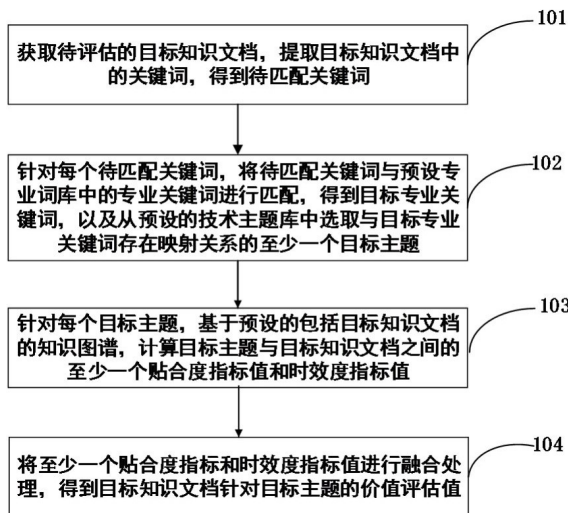
权利要求书3页 说明书13页 附图2页

(54) 发明名称

基于知识图谱的知识文档价值评估方法及装置

(57) 摘要

本公开是关于一种基于知识图谱的知识文档价值评估方法及装置。其中,方法包括:获取待评估的目标知识文档,提取目标知识文档中的关键词,得到待匹配关键词;针对每个待匹配关键词,将待匹配关键词与预设专业词库中的专业关键词进行匹配,得到目标专业关键词,以及从预设的技术主题库中选取与目标专业关键词存在映射关系的至少一个目标主题;针对每个目标主题,基于预设的包括目标知识文档的知识图谱,计算目标主题与目标知识文档之间的至少一个贴合度指标值和时效度指标值;将至少一个贴合度指标值和时效度指标值进行融合处理,得到目标知识文档针对目标主题的价值评估值。本方案提升了知识文档价值评估的准确性。



1. 一种基于知识图谱的知识文档价值评估方法,其特征在于,包括:

获取待评估的目标知识文档,提取所述目标知识文档中的关键词,得到待匹配关键词;

针对每个待匹配关键词,将所述待匹配关键词与预设专业词库中的专业关键词进行匹配,得到目标专业关键词,以及从预设的技术主题库中选取与所述目标专业关键词存在映射关系的至少一个目标主题;

针对每个目标主题,基于预设的包括所述目标知识文档的知识图谱,计算所述目标主题与所述目标知识文档之间的至少一个贴合度指标值和时效度指标值;所述知识图谱以知识文档为节点,以知识文档之间的引用关系为边,所述节点存储有对应的知识文档的属性信息;

将所述至少一个贴合度指标和所述时效度指标值进行融合处理,得到所述目标知识文档针对所述目标主题的价值评估值。

2. 根据权利要求1所述的基于知识图谱的知识文档价值评估方法,其特征在于,在所述获取待评估的目标知识文档之前,方法还包括:

针对所述专业词库中的每个专业关键词,将所述专业关键词分别输入至多个完成训练的大模型,得到每个大模型输出的与所述专业关键词关联的主题,以及所述专业关键词与所述主题之间的初始关联度评分;所述主题为所述技术主题库中的主题;

对所述专业关键词的全部初始关联度评分进行加权求和,得到所述专业关键词与所述主题之间的映射概率;

在所述映射概率大于或者等于预设概率的情况下,确定所述专业关键词与所述主题存在映射关系。

3. 根据权利要求1所述的基于知识图谱的知识文档价值评估方法,其特征在于,所述至少一个贴合度指标值包括以下至少一项:

基于直接引用关系的主题技术引用贴合度;

基于直接被引用关系的主题技术被引用贴合度;

基于树状延伸引用关系的主题技术关联引用贴合度;

基于树状延伸被引用关系的主题技术关联被引用贴合度;

基于文档节点关联引用关系的主题知识文档关联引用贴合度;

基于文档节点关联被引用关系的主题知识文档关联被引用贴合度。

4. 根据权利要求3所述的基于知识图谱的知识文档价值评估方法,其特征在于,所述基于直接引用关系的主题技术引用贴合度通过以下公式计算得到:

$$\text{Theme_quote_fit} = \sum_{A=\text{Theme}(\text{word})} P_keywords(\text{word}, \text{text}) * P_A(\text{word}) * \text{Num}(\text{attribute}[\text{text}, 'quote'])$$

其中, Theme_quote_fit 为主题技术引用贴合度, $P_keywords(\text{word}, \text{text})$ 为关键词概率, $P_A(\text{word})$ 为所述专业词库中待匹配关键词 word 与目标主题 A 之间的映射概率, $\text{Num}(\text{attribute}[\text{text}, 'quote'])$ 为所述知识图谱中以知识文档 text 为节点的直接引用关系 attribute[text, 'quote'] 的数量。

5. 根据权利要求3所述的基于知识图谱的知识文档价值评估方法,其特征在于,所述基于直接被引用关系的主题技术被引用贴合度通过以下公式计算得到:

$$\text{Theme_cited_fit} = \sum_{A=\text{Theme}(\text{word})} P_keywords(\text{word}, \text{text}) * P_A(\text{word}) * \text{Num}(\text{attribute}[\text{text}, 'cited'])$$

其中, Theme_cite_d_fit 为主题技术被引用贴合度, $P_keywords(\text{word}, \text{text})$ 为关键词概率, $P_A(\text{word})$ 为所述专业词库中待匹配关键词 word 与目标主题 A 之间的映射概率, $\text{Num}(\text{attribute}[\text{text}, 'cited'])$ 为为所述知识图谱中以所述目标知识文档 text 为根节点所延伸的树状关系中, 全部被引用关系 $\text{attribute}[\text{text}, 'cited']$ 对应的中间节点和叶子节点的总数量。

6. 根据权利要求3所述的基于知识图谱的知识文档价值评估方法, 其特征在于, 所述基于树状延伸引用关系的主题技术关联引用贴合度通过以下公式计算得到:

$$\text{Theme_quote_association_fit} = \sum_{A=\text{Theme}(\text{word})} P_keywords(\text{word}, \text{text}) * P_A(\text{word}) * \text{Node}(\text{attribute}[\text{text}, 'quote'])$$

其中, $\text{Theme_quote_association_fit}$ 为主题技术关联引用贴合度, $P_keywords(\text{word}, \text{text})$ 为关键词概率, $P_A(\text{word})$ 为所述专业词库中待匹配关键词 word 与目标主题 A 之间的映射概率, $\text{Node}(\text{attribute}[\text{text}, 'quote'])$ 为所述知识图谱中以所述目标知识文档 text 为根节点所延伸的树状关系中, 全部引用关系 $\text{attribute}[\text{text}, 'quote']$ 对应的子节点总数量。

7. 根据权利要求3所述的基于知识图谱的知识文档价值评估方法, 其特征在于, 所述基于树状延伸被引用关系的主题技术关联被引用贴合度通过以下公式计算得到:

$$\text{Theme_cited_association_fit} = \sum_{A=\text{Theme}(\text{word})} P_keywords(\text{word}, \text{text}) * P_A(\text{word}) * \text{Node}(\text{attribute}[\text{text}, 'cited'])$$

其中, $\text{Theme_cited_association_fit}$ 为主题技术关联被引用贴合度, $P_A(\text{word})$ 为所述专业词库中待匹配关键词 word 与目标主题 A 之间的映射概率, $\text{Node}(\text{attribute}[\text{text}, 'cited'])$ 为所述知识图谱中以所述目标知识文档 text 为根节点所延伸的树状关系中, 全部被引用关系 $\text{attribute}[\text{text}, 'cited']$ 对应的叶子节点总数量。

8. 根据权利要求3所述的基于知识图谱的知识文档价值评估方法, 其特征在于, 所述基于文档节点关联引用关系的主题知识文档关联引用贴合度通过以下公式计算得到:

$$\text{Theme_quote_adjacent_fit} = \sum_{A=\text{Theme}(\text{attribute}(\text{text}, 'quote'))} P_build(\text{text}, \text{attribute}(\text{text}, 'quote')) * P_A(\text{word}) * \text{Num}(\text{attribute}[\text{text}, 'quote'])$$

其中, $\text{Theme_quote_adjacent_fit}$ 为主题知识文档关联引用贴合度, $P_build(\text{text}, \text{attribute}(\text{text}, 'quote'))$ 为所述知识图谱中所述目标知识文档与其引用的知识文档之间的关联关系概率, $P_A(\text{word})$ 为所述专业词库中待匹配关键词 word 与目标主题 A 之间的映射概率, $\text{Num}(\text{attribute}[\text{text}, 'quote'])$ 为所述知识图谱中以知识文档 text 为节点的直接引用关系 $\text{attribute}[\text{text}, 'quote']$ 的数量。

9. 根据权利要求3所述的基于知识图谱的知识文档价值评估方法, 其特征在于, 所述基于文档节点关联被引用关系的主题知识文档关联被引用贴合度通过以下公式计算得到:

$$\text{Theme_cited_adjacent_fit} = \sum_{A=\text{Theme}(\text{attribute}(\text{text}, 'cited'))} P_build(\text{text}, \text{attribute}(\text{text}, 'cited')) * P_A(\text{word}) * \text{Num}(\text{attribute}[\text{text}, 'cited'])$$

其中, $\text{Theme_cited_adjacent_fit}$ 为主题知识文档关联被引用贴合度, $P_build(\text{text}, \text{attribute}(\text{text}, 'cited'))$ 为所述知识图谱中所述目标知识文档与引用所述目标知识文档的其他知识文档之间的关联关系概率, $P_A(\text{word})$ 为所述专业词库

中待匹配关键词word与目标主题A之间的映射概率,Num(attribute[text, 'cited'])为所述知识图谱中以所述目标知识文档text为根节点所延伸的树状关系中,全部被引用关系attribute[text, 'cited']对应的中间节点和叶子节点的总数量。

10. 根据权利要求1所述的基于知识图谱的知识文档价值评估方法,其特征在于,所述时效度指标通过以下公式计算得到:

$$\text{Theme_time_validity} = e^{-g(\text{current_time} - \text{attribute}(\text{text}, \text{time}))} * \text{Theme_value} * \sum_{A=\text{Theme}(\text{word})} \text{keywords}(\text{word}, \text{text}) * P_A(\text{word})$$

其中,g为预先设置的文档时效率参数,current_time为当前评估时间点,attribute(text,time)为所述目标知识文档text的存入所述知识图谱的时间,Theme_value为贴合度指标值,keywords(word,text)为待匹配关键词word在目标知识文档text中的重要程度值,P_A(word)为所述专业词库中待匹配关键词word与目标主题A之间的映射概率。

11. 根据权利要求1所述的基于知识图谱的知识文档价值评估方法,其特征在于,所述将所述至少一个贴合度指标和所述时效度指标值进行融合处理,得到所述目标知识文档针对所述目标主题的价值评估值,包括:

获取每个贴合度指标值各自对应的预设权重;

对全部贴合度指标值基于各自的预设权重进行加权求和,得到第一评估值;

计算所述第一评估值与所述时效度指标值之和,得到所述价值评估值。

12. 一种基于知识图谱的知识文档价值评估装置,其特征在于,包括:

提取单元,用于获取待评估的目标知识文档,提取所述目标知识文档中的关键词,得到待匹配关键词;

匹配单元,用于针对每个待匹配关键词,将所述待匹配关键词与预设专业词库中的专业关键词进行匹配,得到目标专业关键词,以及从预设的技术主题库中选取与所述目标专业关键词存在映射关系的至少一个目标主题;

计算单元,用于针对每个目标主题,基于预设的包括所述目标知识文档的知识图谱,计算所述目标主题与所述目标知识文档之间的至少一个贴合度指标值和时效度指标值;所述知识图谱以知识文档为节点,以知识文档之间的引用关系为边,所述节点存储有对应的知识文档的属性信息;

评估单元,用于将所述至少一个贴合度指标和所述时效度指标值进行融合处理,得到所述目标知识文档针对所述目标主题的价值评估值。

基于知识图谱的知识文档价值评估方法及装置

技术领域

[0001] 本公开涉及知识图谱技术领域,尤其涉及一种基于知识图谱的知识文档价值评估方法及装置。

背景技术

[0002] 相关技术中,在当前科研管理与知识服务领域,对知识文档进行客观、精准的科研价值评估是一项重要需求。传统方法主要依赖文献计量学指标,如引用次数、期刊影响因子等,这些指标虽具参考价值,但仅能反映文档的流行度或发表平台的声望,难以深入揭示其内在技术内容的质量与专项贡献。此外,基于简单关键词匹配或固定分类体系的方法,无法动态理解文档技术的深层主题归属,更无法量化该文档在特定技术脉络中的结构性影响。

发明内容

[0003] 为克服相关技术中存在的问题,本公开提供一种基于知识图谱的知识文档价值评估方法及装置。

[0004] 根据本公开实施例的第一方面,提供一种基于知识图谱的知识文档价值评估方法,包括:

[0005] 获取待评估的目标知识文档,提取所述目标知识文档中的关键词,得到待匹配关键词;

针对每个待匹配关键词,将所述待匹配关键词与预设专业词库中的专业关键词进行匹配,得到目标专业关键词,以及从预设的技术主题库中选取与所述目标专业关键词存在映射关系的至少一个目标主题;

针对每个目标主题,基于预设的包括所述目标知识文档的知识图谱,计算所述目标主题与所述目标知识文档之间的至少一个贴合度指标值和时效度指标值;所述知识图谱以知识文档为节点,以知识文档之间的引用关系为边,所述节点存储有对应的知识文档的属性信息;

将所述至少一个贴合度指标和所述时效度指标值进行融合处理,得到所述目标知识文档针对所述目标主题的价值评估值。

[0006] 根据本公开实施例的第二方面,提供一种基于知识图谱的知识文档价值评估装置,包括:

提取单元,用于获取待评估的目标知识文档,提取所述目标知识文档中的关键词,得到待匹配关键词;

匹配单元,用于针对每个待匹配关键词,将所述待匹配关键词与预设专业词库中的专业关键词进行匹配,得到目标专业关键词,以及从预设的技术主题库中选取与所述目标专业关键词存在映射关系的至少一个目标主题;

计算单元,用于针对每个目标主题,基于预设的包括所述目标知识文档的知识图谱,计算所述目标主题与所述目标知识文档之间的至少一个贴合度指标值和时效度指标

值;所述知识图谱以知识文档为节点,以知识文档之间的引用关系为边,所述节点存储有对应的知识文档的属性信息;

评估单元,用于将所述至少一个贴合度指标和所述时效度指标值进行融合处理,得到所述目标知识文档针对所述目标主题的价值评估值。

[0007] 根据本公开实施例的第三方面,一种电子设备,包括:存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时,实现如第一方面中任一项所述的方法。

[0008] 根据本公开实施例的第四方面,提供一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现如第一方面中任一项所述的方法。

[0009] 根据本公开实施例的第五方面,提供一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现如第一方面中任一项所述的方法。

[0010] 本公开的实施例提供的技术方案可以包括以下有益效果:通过提取目标知识文档的关键词并与预设专业词库匹配,能够准确、全面地识别文档所涉及的专业领域,进而通过技术主题库映射确定目标主题,确保了主题识别的专业性和准确性。基于以知识文档为节点、以引用关系为边构建的知识图谱,计算目标主题与文档之间的贴合度指标值,能够有效利用文档间的学术关联网络,量化该文档在特定主题下的学术地位与影响力。同时,通过引入时效度指标值,将知识的新鲜度纳入评估体系,克服了传统方法忽视时间因素的局限。最终,通过融合贴合度与时效度指标,能够从学术关联性和时效性两个维度综合、客观地得出该知识文档在目标主题下的价值评估结果,为科研评价、文献推荐等应用提供了更为科学、全面的决策依据。

[0011] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本公开。

附图说明

[0012] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本发明的实施例,并与说明书一起用于解释本发明的原理。

[0013] 图1是根据一示例性实施例示出的一种基于知识图谱的知识文档价值评估方法的流程图。

[0014] 图2是本公开实施例中提出的专业词库与技术主题库之间的映射关系示意图。

[0015] 图3是根据一示例性实施例示出的一种基于知识图谱的知识文档价值评估装置的框图。

[0016] 图4是根据一示例性实施例示出的一种用于基于知识图谱的知识文档价值评估方法的装置的框图。

具体实施方式

[0017] 这里将详细地对示例性实施例进行说明,其示例表示在附图中。下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本发明相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本发明的一些方面相一致的装置和方法的例子。

[0018] 在本公开实施例使用的术语是仅仅出于描述特定实施例的目的,而非旨在限制本公开实施例。在本公开实施例和所附权利要求书中所使用的单数形式的“一种”和“该”也旨在包括多数形式,除非上下文清楚地表示其他含义。

[0019] 应当理解,尽管在本公开实施例可能采用术语第一、第二、第三等来描述各种信息,但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如,在不脱离本公开实施例范围的情况下,第一信息也可以被称为第二信息,类似地,第二信息也可以被称为第一信息。取决于语境,如在此所使用的词语“如果”及“若”可以被解释成为“在……时”或“当……时”或“响应于确定”。

[0020] 此外,可以使用本公开实施例所示的各种形式的流程,重新排序、增加或删除步骤。例如,本发申请中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行,只要能够实现本公开公开的技术方案所期望的结果,本文在此不进行限制。

[0021] 需要说明的是,本公开的技术方案中,所涉及的用户个人信息的收集、存储、使用、加工、传输、提供和公开等处理,均符合相关法律法规的规定,且不违背公序良俗。

[0022] 图1是根据一示例性实施例示出的一种基于知识图谱的知识文档价值评估方法的流程图,如图1所示,需要说明的是,本公开实施例的基于知识图谱的知识文档价值评估方法应用于基于知识图谱的知识文档价值评估装置中。如图1所示,该方法可以包括以下步骤:

步骤101,获取待评估的目标知识文档,提取目标知识文档中的关键词,得到待匹配关键词。

[0023] 在本公开实施例中,可以通过自然语言处理技术(如文本分析、术语提取算法)从目标知识文档中提取出能够代表其核心内容与技术特征的关键词集合,这些关键词构成了后续主题匹配与关联分析的直接依据,即待匹配关键词。

[0024] 在一个实施例中,可以采用完成训练的自然语言处理模型提取目标知识文档中的关键词。

[0025] 在本公开一些实施例中,在步骤101之前,方法还可以包括:

步骤a1,针对专业词库中的每个专业关键词,将专业关键词分别输入至多个完成训练的大模型,得到每个大模型输出的与专业关键词关联的主题,以及专业关键词与主题之间的初始关联度评分。

[0026] 其中,主题为技术主题库中的主题。

[0027] 需要说明的是,专业词库可以是一个经过系统化构建与维护的、覆盖特定领域或学科的专业术语与关键概念的标准化词集合,其内部存储有大量具有明确专业指向性的关键词(如技术术语、核心概念、特定表述等)。专业词库不仅作为关键词提取与匹配的基准参照,还为后续基于大模型的语义映射提供原始输入,确保能够准确识别文档中的专业内容并将其关联至相应的技术主题,从而为整个知识文档价值评估体系奠定规范、可靠的语义基础。

[0028] 此外,技术主题库是一个结构化、层次化的主题分类体系,用于系统化地组织与管理特定技术领域或学科方向下的核心研究主题与方向。技术主题库不仅作为知识内容分类与归类标准框架,还通过预设的映射规则与专业词库中的关键词建立语义关联,从而在知识文档价值评估过程中作为主题识别、内容分析与价值量化的核心参照系,最终支撑系

统实现基于主题的知识聚合、关联分析与价值判断。

[0029] 在本公开实施例中,针对专业词库中的每个专业关键词,将其同时输入至多个已完成训练的通用AI大模型,利用各模型对词语的深层语义解析与上下文关联能力,分别输出该关键词所关联的一个或多个技术主题(主题均来源于预设的技术主题库),并给出对应的初始关联度评分。从而有效融合了不同模型的认知优势,为后续构建细粒度、概率化的关键词-主题映射提供了可靠且多维度的数据基础,从而支撑整个评估系统实现更精准的主题识别与内容关联分析。

[0030] 步骤a2,对专业关键词的全部初始关联度评分进行加权求和,得到专业关键词与主题之间的映射概率。

[0031] 在本公开实施例中,针对每个专业关键词在不同大模型中获得的多个初始关联度评分,根据各模型预设的权重系数(通常基于模型历史表现或置信度动态设定),对同一主题对应的全部初始关联度评分进行加权融合计算,最终生成该关键词与特定技术主题之间唯一的映射概率值。这一过程不仅整合了多模型的语义判断优势,而且通过权值分配增强了结果的鲁棒性与可信度,为后续知识图谱构建和价值评估提供了归一化、可比较的概率数据基础。

[0032] 在一个实施例中,可以采用以下公式计算专业关键词与主题之间的映射概率 $P_{\text{map}}(A(k) | B_n)$:

$$\text{Theme}(\text{API}(i), B_n) = \{A(k)\}$$

[0033] $\text{Score}_{\text{init}}(\text{Theme}(\text{API}(i), B_n)) = \{\text{Score}_{\text{init}}(k)\}$

[0034] $\{\text{Score}(k)\} = \text{Softmax}(\{\text{Score}_{\text{init}}(k)\})$

[0035] 其中, $\text{Theme}(\text{API}(i), B_n)$ 为第*i*个大模型API针对关键词 B_n 返回的关联技术主题集合,记为 $\{A(k)\}$, $\text{Score}_{\text{init}}(k)$ 为第*i*个大模型接口API返回的、归一化前的原始关联度评分, $A(k)$ 为技术主题库中的第*k*个技术主题, $\text{Score}(k)$ 为对初始评分进行 Softmax 归一化处理后得到的评分

为每个API(*i*)分配了权重 γ_i (全部 γ 之和为1), 关键词 B_n 与主题 $A(k)$ 的最终映射概率,通过对所有API的归一化评分进行加权求和得到:

$$P_{\text{map}}(A(k) | B_n) = \sum_i [\gamma_i \cdot \text{Score}_i(k)]$$

[0036] 其中, $\text{Score}(k) \in (0, 1)$ 且全部 $\text{Score}(k)$ 之和为1, $\text{Score}_i(k)$ 为API(*i*)对主题 $A(k)$ 返回的归一化评分。

[0037] 步骤a3,在映射概率大于或者等于预设概率的情况下,确定专业关键词与主题存在映射关系。

[0038] 在一个实施例中,在完成多模型加权融合得到关键词与主题之间的映射概率后,将映射概率与一个预设的阈值进行比较;当映射概率大于或等于该阈值时,正式确认该专业关键词与该技术主题之间存在有效的映射关系,并将其纳入后续的知识图谱构建与计算体系。这一机制通过概率过滤有效剔除了关联度较弱或置信度不足的模糊匹配,显著提升了主题归类的准确性与知识图谱数据的可靠性,从而为后续的价值评估奠定了高质量的结构化基础。

[0039] 在一个示例中,如图2所示,展示了本技术方案中预设的技术主题库与专业词库之间的对应关系结构:技术主题库包含了从主题A1到主题An的技术主题;右侧的专业词库则包含了从关键字B1到关键字Bn的规范化关键词集合;图中两侧元素之间的连线表示了专业词库中的每个关键字与技术主题库中的一个或多个主题之间存在预设的映射关系。

[0040] 步骤102,针对每个待匹配关键词,将待匹配关键词与预设专业词库中的专业关键词进行匹配,得到目标专业关键词,以及从预设的技术主题库中选取与目标专业关键词存在映射关系的至少一个目标主题。

[0041] 在一个实施例中,可以将待匹配关键词与预设的专业词库中的专业关键词进行比对与匹配,与待匹配关键词相同或者相似度高于预设相似度的目标专业关键词,基于已建立的关键词-主题映射关系,从技术主题库中自动选取与目标专业关键词相关联的一个或多个“目标主题”。此过程实现了从文档原始关键词到规范化技术主题的精准转换,为后续基于主题的知识图谱分析与价值量化提供了明确的评估对象与维度。

[0042] 在一些实施例中,若待匹配关键词与预设的专业词库中的专业关键词匹配失败,可以将待匹配关键词添加至专业词库中,采用本公开步骤a1-a3中公开的将关键词与主题进行映射的具体实现方式,构建待匹配关键词与至少一个主题之间的映射关系,在完成映射关系构建之后,执行步骤102。

[0043] 步骤103,针对每个目标主题,基于预设的包括目标知识文档的知识图谱,计算目标主题与目标知识文档之间的至少一个贴合度指标值和时效度指标值。

[0044] 其中,知识图谱以知识文档为节点,以知识文档之间的引用关系为边,节点存储有对应的知识文档的属性信息。

[0045] 在一个实施例中,时效度指标值可以是一个用于量化知识文档价值随时间动态衰减的关键参数,基于文档的发表时间或最新更新时间与当前评估时间点的时间差,通过预设的衰减函数(如指数衰减函数)进行计算,其中系统可调参数来控制衰减速率。该指标值直观反映了知识的新鲜度与时效性,其数值通常介于0到1之间,越新的文档其值越接近1,越旧的文档则越接近0。时效度指标与基于知识图谱结构计算的静态贴合度指标相乘或加权融合,从而将时间维度纳入评估体系,确保能够优先识别并赋予近期高关联文档以更高的综合价值评分,有效模拟了知识价值随技术发展自然演化的现实规律。

[0046] 在一个实施例中,知识图谱可以是一种以结构化语义网络形式组织知识文档及其关联关系的核心数据模型,以知识文档为节点、以文档间的引用关系为边构建网络拓扑,并在节点中存储文档的属性信息(如关键词、发表时间、主题归属概率等)。该图谱不仅静态地描述了文档间的学术关联结构,更通过节点属性中预计算的概率化数据(如文档属于某主题的概率)动态支持多维度贴合度指标的量化计算,从而为综合评估文档在特定主题下的学术影响力与时效性价值提供了可计算、可解释的网络化数据基础。

[0047] 在本公开实施例中,利用预先构建的知识图谱,对目标知识文档与特定技术主题之间的关联强度进行多维度、结构化的计算,基于图谱的拓扑连接与节点属性,计算一系列反映文档在该主题下学术影响力的贴合度指标值(例如直接与间接的引用/被引用关联强度);同时,结合文档自身的时间属性,计算其时效度指标值,实现了从静态的主题匹配到动态的、基于网络结构与时间效应的综合量化评估,为最终的价值判定提供了核心数据输入。

[0048] 在本公开一些实施例中,至少一个贴合度指标值包括以下至少一项:

基于直接引用关系的主题技术引用贴合度；
 基于直接被引用关系的主题技术被引用贴合度；
 基于树状延伸引用关系的主题技术关联引用贴合度；
 基于树状延伸被引用关系的主题技术关联被引用贴合度；
 基于文档节点关联引用关系的主题知识文档关联引用贴合度；
 基于文档节点关联被引用关系的主题知识文档关联被引用贴合度。

[0049] 在本公开一些实施例中,基于直接引用关系的主题技术引用贴合度通过以下公式计算得到:

$$\text{Theme_quote_fit} = \sum_{A=\text{Theme}(\text{word})} P_keywords(\text{word}, \text{text}) * P_A(\text{word}) * \text{Num}(\text{attribute}[\text{text}, 'quote'])$$

[0050] 其中, Theme_quote_fit 为主题技术引用贴合度, $P_keywords(\text{word}, \text{text})$ 为知识图谱中待匹配关键词针对目标知识文档的关键词概率, $P_A(\text{word})$ 为专业词库中待匹配关键词 word 与目标主题 A 之间的映射概率, $\text{Num}(\text{attribute}[\text{text}, 'quote'])$ 为知识图谱中以知识文档 text 为节点的直接引用关系 $\text{attribute}[\text{text}, 'quote']$ 的数量。

[0051] 在本公开实施例中,将关键词在文档中的重要性 ($P_keywords$) 与关键词和主题的语义映射概率 (P_A) 相乘,能够准确反映文档内容与该主题的语义贴合强度,克服了传统方法仅依赖词频或简单匹配的不足;进一步引入基于知识文档在知识图谱中的直接引用关系数量 (Num),将文档的学术影响力纳入计算,使得评估结果既包含内容相关性、又涵盖学术认可度;最终通过加权求和方式,综合语义与结构两类信息,输出具有良好可解释性和可比性的贴合度指标,为科研评价、文献推荐等场景提供更科学、更全面的决策依据。

[0052] 在本公开一些实施例中,基于直接被引用关系的主题技术被引用贴合度通过以下公式计算得到:

$$\text{Theme_cited_fit} = \sum_{A=\text{Theme}(\text{word})} P_keywords(\text{word}, \text{text}) * P_A(\text{word}) * \text{Num}(\text{attribute}[\text{text}, 'cited'])$$

[0053] 其中, Theme_cited_fit 为主题技术被引用贴合度, $P_keywords(\text{word}, \text{text})$ 为知识图谱中待匹配关键词针对目标知识文档的关键词概率, $P_A(\text{word})$ 为专业词库中待匹配关键词 word 与目标主题 A 之间的映射概率, $\text{Num}(\text{attribute}[\text{text}, 'cited'])$ 为知识图谱中以目标知识文档 text 为根节点所延伸的树状关系中,全部被引用关系 $\text{attribute}[\text{text}, 'cited']$ 对应的中间节点和叶子节点的总数量。

[0054] 在本公开实施例中,以关键词概率和主题映射概率的乘积精准捕捉文档内容与目标主题的语义关联强度,确保评估建立在扎实的内容分析基础上,进一步引入以文档为根节点的被引用关系树状网络中全部节点的数量,不仅统计直接被引次数,更涵盖多级间接引用,从而全面衡量文档在学术传承网络中的持久影响力与知识扩散广度;最终将语义相关性、与结构影响力有机结合,克服了传统方法仅依赖直接引用次数或单纯内容匹配的局限,为识别具有长期、广泛学术价值的核心文献提供了更科学、更具解释力的评估依据。

[0055] 在本公开一些实施例中,基于树状延伸引用关系的主题技术关联引用贴合度通过以下公式计算得到:

$$\text{Theme_quote_association_fit} = \sum_{A=\text{Theme}(\text{word})} P_keywords(\text{word}, \text{text}) * P_A(\text{word}) * \text{Node}(\text{attribute}[\text{text}, 'quote'])$$

[0056] 其中, $\text{Theme_quote_association_fit}$ 为主题技术关联引用贴合度, $P_keywords(word, text)$ 为知识图谱中待匹配关键词针对目标知识文档的关键词概率, $P_A(word)$ 为专业词库中待匹配关键词 $word$ 与目标主题 A 之间的映射概率, $\text{Node}(\text{attribute}[text, 'quote'])$ 为知识图谱中以目标知识文档 $text$ 为根节点所延伸的树状关系中, 全部引用关系 $\text{attribute}[text, 'quote']$ 对应的子节点总数量。

[0057] 可以理解的是, $\text{attribute}[text, 'quote']$ 可以是指从目标知识文档对应的节点 $text$ 出发, 指向其所引用的其他文档的所有引用关系边的集合。

[0058] 在本公开实施例中, 以关键词与主题的映射概率加权关键词重要性的方式, 精准量化文档内容与目标主题的语义相关性; 进一步引入以文档为根节点的引用关系树中全部子节点的数量, 不仅统计直接引用, 更涵盖多级间接引用所形成的影响链条, 从而全面反映该文档在学术网络中的知识扩散深度与传承范围; 最终将内容相关性与多层次结构影响力相结合, 克服了传统评估方法仅关注直接引用或静态关联的局限, 为识别在特定技术主题下具有持续、深远学术价值的核心文献提供了更立体、更科学的量化依据。

[0059] 在本公开一些实施例中, 基于树状延伸被引用关系的主题技术关联被引用贴合度通过以下公式计算得到:

$$\text{Theme_cited_association_fit} = \sum_{A=\text{Theme}(word)} P_keywords(word, text) * P_A(word) * \text{Node}(\text{attribute}[text, 'cited'])$$

[0060] 其中, $\text{Theme_cited_association_fit}$ 为主题技术关联被引用贴合度, $P_A(word)$ 为专业词库中待匹配关键词 $word$ 与目标主题 A 之间的映射概率, $\text{Node}(\text{attribute}[text, 'cited'])$ 为知识图谱中以目标知识文档 $text$ 为根节点所延伸的树状关系中, 全部被引用关系 $\text{attribute}[text, 'cited']$ 对应的叶子节点总数量。

[0061] 可以理解的是, $\text{attribute}[text, 'cited']$ 可以是指从其他文档节点指向目标文档节点 $text$ 的所有被引用关系边的集合。

[0062] 在本公开实施例中, 基于关键词与主题的映射概率及关键词重要性, 准确捕捉文档内容与目标主题的语义关联强度; 通过统计以文档为根节点的被引用关系树中全部叶子节点的数量, 不仅涵盖直接被引, 更聚焦于引用链末端的影响范围, 从而有效反映该文档在学术传承中的最终影响力和知识扩散的终端覆盖广度; 最终将语义相关性与深层次结构影响力相结合, 克服了传统方法仅关注直接被引次数或忽视间接影响的局限, 为识别在特定技术主题下具有长期、稳定且广泛学术价值的核心文献提供了更具深度与结构感知能力的评估依据。

[0063] 在本公开一些实施例中, 基于文档节点关联引用关系的主题知识文档关联引用贴合度通过以下公式计算得到:

$$\text{Theme_quote_adjacent_fit} = \sum_{A=\text{Theme}(\text{attribute}[text, 'quote'])} P_build(text, \text{attribute}[text, 'quote']) * P_A(word) * \text{Num}(\text{attribute}[text, 'quote'])$$

[0064] 其中, $\text{Theme_quote_adjacent_fit}$ 为主题知识文档关联引用贴合度, $P_build(text, \text{attribute}[text, 'quote'])$ 为知识图谱中目标知识文档与其引用的知识文档之间的关联关系概率, $P_A(word)$ 为专业词库中待匹配关键词 $word$ 与目标主题 A 之间的映射概率, $\text{Num}(\text{attribute}[text, 'quote'])$ 为知识图谱中以知识文档 $text$ 为节点的直接引用关系 $\text{attribute}[text, 'quote']$ 的数量。

[0065] 作为一种可能实施方式的示例,知识图谱中目标知识文档与其引用的知识文档之间的关联关系概率的确定方式可以为:基于文档内容的语义相似度(如通过预训练语言模型提取文本向量后计算余弦相似度),还可以基于共同引用的主题或关键词权重(如利用已建立的关键词-主题映射概率进行联合推断),还可以基于引用网络的图结构度量(如利用图嵌入算法学习节点表示后计算关联强度),或融合以上多种特征通过机器学习模型(如逻辑回归、神经网络)进行训练预测。该关联关系概率值动态反映了文档间引用关系在语义、主题与结构层面的综合关联强度,为后续贴合度计算提供了细粒度、可量化的关系权重。

[0066] 在本公开实施例中,将知识图谱中预计算的文档间关联关系概率(P_{build})、关键词与目标主题的语义映射概率(P_A)以及直接引用关系的数量(Num)三者相乘并求和,不仅量化了引用行为的广度,更通过前两个概率因子深度刻画了每一次引用在语义与主题层面的相关性与强度,克服了传统引用分析中仅统计引用次数而忽略引用内容相关性与关系质量的局限,使得评估结果能够更准确地区分出那些在特定技术主题下既积极引用他人成果、且所引内容与本主题高度相关的知识文档,从而为识别具有高质量学术关联与知识整合能力的文献提供了更科学、更具解释力的量化依据。

[0067] 在本公开一些实施例中,基于文档节点关联被引用关系的主题知识文档关联被引用贴合度通过以下公式计算得到:

$$\text{Theme_cited_adjacent_fit} = \sum_{A=\text{Theme}(\text{attribute}(\text{text}, \text{'cited'})}) P_{\text{build}}(\text{text}, \text{attribute}(\text{text}, \text{'cited'})) * P_A(\text{word}) * \text{Num}(\text{attribute}[\text{text}, \text{'cited'}])$$

[0068] 其中, $\text{Theme_cited_adjacent_fit}$ 为主题知识文档关联被引用贴合度, $P_{\text{build}}(\text{text}, \text{attribute}(\text{text}, \text{'cited'}))$ 为知识图谱中目标知识文档与引用目标知识文档的其他知识文档之间的关联关系概率, $P_A(\text{word})$ 为专业词库中待匹配关键词 word 与目标主题 A 之间的映射概率, $\text{Num}(\text{attribute}[\text{text}, \text{'cited'}])$ 为知识图谱中以目标知识文档 text 为根节点所延伸的树状关系中,全部被引用关系 $\text{attribute}[\text{text}, \text{'cited'}]$ 对应的中间节点和叶子节点的总数量。

[0069] 在本公开实施例中,分子中的三项乘积(文档间关联关系概率 $\times$ 关键词-主题映射概率 $\times$ 被引用网络节点总数)创造性地将关系强度、语义相关性 with 影响广度三个维度进行耦合计算。这使得评估结果既能通过 P_{build} 区分不同引用来源的重要性,又能通过 P_A 确保评估聚焦于目标主题,还能通过 Num 客观反映该文档在学术网络中的总体影响规模。该公式避免了简单加总的粗糙性,提供了一种更精细、更综合的量化方法,用以识别那些在特定技术领域不仅被广泛引用,且这些引用大多来自高相关文档的高价值核心文献。

[0070] 在本公开一些实施例中,时效度指标通过以下公式计算得到:

$$\text{Theme_time_validity} = e^{-g(\text{current_time} - \text{attribute}(\text{text}, \text{time}))} * \text{Theme_value} * \sum_{A=\text{Theme}(\text{word})} \text{keywords}(\text{word}, \text{text}) * P_A(\text{word})$$

[0071] 其中, g 为预先设置的文档时效率参数, current_time 为当前评估时间点, $\text{attribute}(\text{text}, \text{time})$ 为目标知识文档 text 的存入知识图谱的时间, Theme_value 为综合贴合度指标值, $\text{keywords}(\text{word}, \text{text})$ 为待匹配关键词 word 在目标知识文档 text 中的重要程度值, $P_A(\text{word})$ 为专业词库中待匹配关键词 word 与目标主题 A 之间的映射概率。

[0072] 在一个实施例中,综合贴合度指标值 Theme_value 采用以下公式计算得到:

$$\text{Theme_value} = \sum_{i=1}^6 \mu_i * t_i$$

[0073] 其中, t_i 为第 i 个贴合度指标值, μ_i 为第 i 个贴合度指标值的预设权重。

[0074] 在本公开实施例中, 通过将指数时间衰减、综合贴合度指标值与映射概率三者耦合, 实现了对知识文档价值的时效修正。该指标不仅通过可调节的衰减速率精确模拟了知识随技术发展自然老化的规律, 克服了传统静态评估体系忽视时间因素的固有缺陷。

[0075] 步骤104, 将至少一个贴合度指标和时效度指标值进行融合处理, 得到目标知识文档针对目标主题的价值评估值。

[0076] 在本公开一些实施例中, 步骤104具体可以包括以下步骤:

获取每个贴合度指标值各自对应的预设权重;

对全部贴合度指标值基于各自的预设权重进行加权求和, 得到第一评估值;

计算第一评估值与时效度指标值之和, 得到价值评估值。

[0077] 在本公开实施例中, 通过加权求和将多个反映文档结构与主题关联的指标聚合成第一评估值, 将该静态评估值与时效度指标值相加, 从而在最终的价值评估结果中同时嵌入文档的影响力与时效性贡献, 确保输出结果兼具学术网络重要性与时间相关性, 为后续的排序、推荐或决策提供直观且全面的量化依据。

[0078] 根据本公开实施例提出的基于知识图谱的知识文档价值评估方法, 通过提取目标知识文档的关键词并与预设专业词库匹配, 能够准确、全面地识别文档所涉及的专业领域, 进而通过技术主题库映射确定目标主题, 确保了主题识别的专业性和准确性。基于以知识文档为节点、以引用关系为边构建的知识图谱, 计算目标主题与文档之间的贴合度指标值, 能够有效利用文档间的学术关联网络, 量化该文档在特定主题下的学术地位与影响力。同时, 通过引入时效度指标值, 将知识的新鲜度纳入评估体系, 克服了传统方法忽视时间因素的局限。最终, 通过融合贴合度与时效度指标, 能够从学术关联性和时效性两个维度综合、客观地得出该知识文档在目标主题下的价值评估结果, 为科研评价、文献推荐等应用提供了更为科学、全面的决策依据。

[0079] 图3是根据一示例性实施例示出的一种基于知识图谱的知识文档价值评估装置框图。参照图3, 该装置包括提取单元301, 匹配单元302, 计算单元303和评估单元304。

[0080] 其中, 提取单元301, 用于获取待评估的目标知识文档, 提取目标知识文档中的关键词, 得到待匹配关键词;

匹配单元302, 用于针对每个待匹配关键词, 将待匹配关键词与预设专业词库中的专业关键词进行匹配, 得到目标专业关键词, 以及从预设的技术主题库中选取与目标专业关键词存在映射关系的至少一个目标主题;

计算单元303, 用于针对每个目标主题, 基于预设的包括目标知识文档的知识图谱, 计算目标主题与目标知识文档之间的至少一个贴合度指标值和时效度指标值; 知识图谱以知识文档为节点, 以知识文档之间的引用关系为边, 节点存储有对应的知识文档的属性信息;

评估单元304, 用于将至少一个贴合度指标和时效度指标值进行融合处理, 得到目标知识文档针对目标主题的价值评估值。

[0081] 在本公开一些实施例中, 装置还可以包括:

主题确定单元, 用于针对专业词库中的每个专业关键词, 将专业关键词分别输入至多个完成训练的大模型, 得到每个大模型输出的与专业关键词关联的主题, 以及专业关

关键词与主题之间的初始关联度评分；主题为技术主题库中的主题；

求和单元,用于对专业关键词的全部初始关联度评分进行加权求和,得到专业关键词与主题之间的映射概率；

确定单元,用于在映射概率大于或者等于预设概率的情况下,确定专业关键词与主题存在映射关系。

[0082] 在本公开一些实施例中,至少一个贴合度指标值包括以下至少一项：

基于直接引用关系的主题技术引用贴合度；

基于直接被引用关系的主题技术被引用贴合度；

基于树状延伸引用关系的主题技术关联引用贴合度；

基于树状延伸被引用关系的主题技术关联被引用贴合度；

基于文档节点关联引用关系的主题知识文档关联引用贴合度；

基于文档节点关联被引用关系的主题知识文档关联被引用贴合度。

[0083] 在本公开一些实施例中,基于直接引用关系的主题技术引用贴合度通过以下公式计算得到：

$$\text{Theme_quote_fit} = \sum_{A=\text{Theme}(\text{word})} P_keywords(\text{word}, \text{text}) * P_A(\text{word}) * \text{Num}(\text{attribute}[\text{text}, 'quote'])$$

[0084] 其中, Theme_quote_fit 为主题技术引用贴合度, $P_keywords(\text{word}, \text{text})$ 为关键词概率, $P_A(\text{word})$ 为专业词库中待匹配关键词 word 与目标主题 A 之间的映射概率, $\text{Num}(\text{attribute}[\text{text}, 'quote'])$ 为知识图谱中以知识文档 text 为节点的直接引用关系 attribute[text, 'quote'] 的数量。

[0085] 在本公开一些实施例中,基于直接被引用关系的主题技术被引用贴合度通过以下公式计算得到：

$$\text{Theme_cited_fit} = \sum_{A=\text{Theme}(\text{word})} P_keywords(\text{word}, \text{text}) * P_A(\text{word}) * \text{Num}(\text{attribute}[\text{text}, 'cited'])$$

[0086] 其中, Theme_cited_fit 为主题技术被引用贴合度, $P_keywords(\text{word}, \text{text})$ 为关键词概率, $P_A(\text{word})$ 为专业词库中待匹配关键词 word 与目标主题 A 之间的映射概率, $\text{Num}(\text{attribute}[\text{text}, 'cited'])$ 为为知识图谱中以目标知识文档 text 为根节点所延伸的树状关系中,全部被引用关系 attribute[text, 'cited'] 对应的中间节点和叶子节点的总数量。

[0087] 在本公开一些实施例中,基于树状延伸引用关系的主题技术关联引用贴合度通过以下公式计算得到：

$$\text{Theme_quote_association_fit} = \sum_{A=\text{Theme}(\text{word})} P_keywords(\text{word}, \text{text}) * P_A(\text{word}) * \text{Node}(\text{attribute}[\text{text}, 'quote'])$$

[0088] 其中, Theme_quote_association_fit 为主题技术关联引用贴合度, $P_keywords(\text{word}, \text{text})$ 为关键词概率, $P_A(\text{word})$ 为专业词库中待匹配关键词 word 与目标主题 A 之间的映射概率, $\text{Node}(\text{attribute}[\text{text}, 'quote'])$ 为知识图谱中以目标知识文档 text 为根节点所延伸的树状关系中,全部引用关系 attribute[text, 'quote'] 对应的子节点总数量。

[0089] 在本公开一些实施例中,基于树状延伸被引用关系的主题技术关联被引用贴合度

通过以下公式计算得到：

$$\text{Theme_cited_association_fit} = \sum_{A=\text{Theme}(\text{word})} P_keywords(\text{word}, \text{text}) * P_A(\text{word}) * \text{Node}(\text{attribute}[\text{text}, 'cited'])$$

[0090] 其中, Theme_cited_association_fit 为主题技术关联被引用贴合度, $P_A(\text{word})$ 为专业词库中待匹配关键词word与目标主题A之间的映射概率, Node(attribute[text, 'cited']) 为知识图谱中以目标知识文档text为根节点所延伸的树状关系中, 全部被引用关系attribute[text, 'cited']对应的叶子节点总数量。

[0091] 在本公开一些实施例中, 基于文档节点关联引用关系的主题知识文档关联引用贴合度通过以下公式计算得到：

$$\text{Theme_quote_adjacent_fit} = \sum_{A=\text{Theme}(\text{attribute}[\text{text}, 'quote'])} P_build(\text{text}, \text{attribute}[\text{text}, 'quote']) * P_A(\text{word}) * \text{Num}(\text{attribute}[\text{text}, 'quote'])$$

[0092] 其中, Theme_quote_adjacent_fit 为主题知识文档关联引用贴合度, $P_build(\text{text}, \text{attribute}[\text{text}, 'quote'])$ 为知识图谱中目标知识文档与其引用的知识文档之间的关联关系概率, $P_A(\text{word})$ 为专业词库中待匹配关键词word与目标主题A之间的映射概率, Num(attribute[text, 'quote']) 为知识图谱中以知识文档text为节点的直接引用关系attribute[text, 'quote']的数量。

[0093] 在本公开一些实施例中, 基于文档节点关联被引用关系的主题知识文档关联被引用贴合度通过以下公式计算得到：

$$\text{Theme_cited_adjacent_fit} = \sum_{A=\text{Theme}(\text{attribute}[\text{text}, 'cited'])} P_build(\text{text}, \text{attribute}[\text{text}, 'cited']) * P_A(\text{word}) * \text{Num}(\text{attribute}[\text{text}, 'cited'])$$

[0094] 其中, Theme_cited_adjacent_fit 为主题知识文档关联被引用贴合度, $P_build(\text{text}, \text{attribute}[\text{text}, 'cited'])$ 为知识图谱中目标知识文档与引用目标知识文档的其他知识文档之间的关联关系概率, $P_A(\text{word})$ 为专业词库中待匹配关键词word与目标主题A之间的映射概率, Num(attribute[text, 'cited']) 为知识图谱中以目标知识文档text为根节点所延伸的树状关系中, 全部被引用关系attribute[text, 'cited']对应的中间节点和叶子节点的总数量。

[0095] 在本公开一些实施例中, 时效度指标通过以下公式计算得到：

$$\text{Theme_time_validity} = e^{-g(\text{current_time} - \text{attribute}[\text{text}, \text{time}])} * \text{Theme_value} * \sum_{A=\text{Theme}(\text{word})} keywords(\text{word}, \text{text}) * P_A(\text{word})$$

[0096] 其中, g为预先设置的文档时效率参数, current_time为当前评估时间点, attribute(text, time) 为目标知识文档text的存入知识图谱的时间, Theme_value为贴合度指标值, keywords(word, text) 为待匹配关键词word在目标知识文档text中的重要程度值, $P_A(\text{word})$ 为专业词库中待匹配关键词word与目标主题A之间的映射概率。

[0097] 在本公开一些实施例中, 评估单元304具体可以用于：

获取每个贴合度指标值各自对应的预设权重；

对全部贴合度指标值基于各自的预设权重进行加权求和, 得到第一评估值；

计算第一评估值与时效度指标值之和, 得到价值评估值。

[0098] 关于上述实施例中的装置, 其中各个模块执行操作的具体方式已经在有关该方法的实施例中进行了详细描述, 此处将不做详细阐述说明。

[0099] 根据本公开实施例提出的基于知识图谱的知识文档价值评估装置, 通过提取目标知识文档的关键词并与预设专业词库匹配, 能够准确、全面地识别文档所涉及的专业领域,

进而通过技术主题库映射确定目标主题,确保了主题识别的专业性和准确性。基于以知识文档为节点、以引用关系为边构建的知识图谱,计算目标主题与文档之间的贴合度指标值,能够有效利用文档间的学术关联网络,量化该文档在特定主题下的学术地位与影响力。同时,通过引入时效度指标值,将知识的新鲜度纳入评估体系,克服了传统方法忽视时间因素的局限。最终,通过融合贴合度与时效度指标,能够从学术关联性和时效性两个维度综合、客观地得出该知识文档在目标主题下的价值评估结果,为科研评价、文献推荐等应用提供了更为科学、全面的决策依据。

[0100] 图4是根据一示例性实施例示出的一种用于基于知识图谱的知识文档价值评估方法的装置的框图。例如,装置400可以是电子设备,例如可以是移动电话,计算机,数字广播终端,消息收发设备,平板设备,个人数字助理等。

[0101] 参照图4,装置400可以包括以下一个或多个组件:处理组件402,存储器404,电力组件406,多媒体组件408,音频组件410,输入/输出(I/O)的接口412,传感器组件414,以及通信组件416。

[0102] 处理组件402通常控制装置400的整体操作,诸如与显示,电话呼叫,数据通信,相机操作和记录操作相关联的操作。处理组件402可以包括一个或多个处理器420来执行指令,以完成上述的方法的全部或部分步骤。此外,处理组件402可以包括一个或多个模块,便于处理组件402和其他组件之间的交互。例如,处理组件402可以包括多媒体模块,以方便多媒体组件408和处理组件402之间的交互。

[0103] 存储器404被配置为存储各种类型的数据以支持在设备400的操作。这些数据的示例包括用于在装置400上操作的任何应用程序或方法的指令,联系人数据,电话簿数据,消息,图片,视频等。存储器404可以由任何类型的易失性或非易失性存储设备或者它们的组合实现,如静态随机存取存储器(SRAM),电可擦除可编程只读存储器(EEPROM),可擦除可编程只读存储器(EPROM),可编程只读存储器(PROM),只读存储器(ROM),磁存储器,快闪存储器,磁盘或光盘。

[0104] 电力组件406为装置400的各种组件提供电力。电力组件406可以包括电源管理系统,一个或多个电源,及其他与为装置400生成、管理和分配电力相关联的组件。

[0105] 多媒体组件408包括在所述装置400和用户之间的提供一个输出接口的屏幕。在一些实施例中,屏幕可以包括液晶显示器(LCD)和触摸面板(TP)。如果屏幕包括触摸面板,屏幕可以被实现为触摸屏,以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界,而且还检测与所述触摸或滑动操作相关的持续时间和压力。在一些实施例中,多媒体组件408包括一个前置摄像头和/或后置摄像头。当设备400处于操作模式,如拍摄模式或视频模式时,前置摄像头和/或后置摄像头可以接收外部的多媒体数据。每个前置摄像头和后置摄像头可以是一个固定的光学透镜系统或具有焦距和光学变焦能力。

[0106] 音频组件410被配置为输出和/或输入音频信号。例如,音频组件410包括一个麦克风(MIC),当装置400处于操作模式,如呼叫模式、记录模式和语音识别模式时,麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器404或经由通信组件416发送。在一些实施例中,音频组件410还包括一个扬声器,用于输出音频信号。

[0107] I/O接口412为处理组件402和外围接口模块之间提供接口,上述外围接口模块可

以是键盘,点击轮,按钮等。这些按钮可包括但不限于:主页按钮、音量按钮、启动按钮和锁定按钮。

[0108] 传感器组件414包括一个或多个传感器,用于为装置400提供各个方面的状态评估。例如,传感器组件414可以检测到设备400的打开/关闭状态,组件的相对定位,例如所述组件为装置400的显示器和小键盘,传感器组件414还可以检测装置400或装置400一个组件的位置改变,用户与装置400接触的存在或不存在,装置400方位或加速/减速和装置400的温度变化。传感器组件414可以包括接近传感器,被配置用来在没有任何的物理接触时检测附近物体的存在。传感器组件414还可以包括光传感器,如CMOS或CCD图像传感器,用于在成像应用中使用。在一些实施例中,该传感器组件414还可以包括加速度传感器,陀螺仪传感器,磁传感器,压力传感器或温度传感器。

[0109] 通信组件416被配置为便于装置400和其他设备之间有线或无线方式的通信。装置400可以接入基于通信标准的无线网络,如WiFi,2G或3G,或它们的组合。在一个示例性实施例中,通信组件416经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中,所述通信组件416还包括近场通信(NFC)模块,以促进短程通信。例如,在NFC模块可基于射频识别(RFID)技术,红外数据协会(IrDA)技术,超宽带(UWB)技术,蓝牙(BT)技术和其他技术来实现。

[0110] 在示例性实施例中,装置400可以被一个或多个应用专用集成电路(ASIC)、数字信号处理器(DSP)、数字信号处理设备(DSPD)、可编程逻辑器件(PLD)、现场可编程门阵列(FPGA)、控制器、微控制器、微处理器或其他电子元件实现,用于执行上述方法。

[0111] 在示例性实施例中,还提供了一种包括指令的非临时性计算机可读存储介质,例如包括指令的存储器404,上述指令可由装置400的处理器420执行以完成上述方法。例如,所述非临时性计算机可读存储介质可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0112] 在示例性实施例中,还提供了一种计算机程序产品,包括计算机程序,所述计算机程序在被装置400的处理器420执行时实现上述方法。

[0113] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本发明的其它实施方案。本公开旨在涵盖本发明的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本发明的一般性原理并包括本公开未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本发明的真正范围和精神由下面的权利要求指出。

[0114] 应当理解的是,本发明并不局限于上面已经描述并在附图中示出的精确结构,并且可以在不脱离其范围进行各种修改和改变。本发明的范围仅由所附的权利要求来限制。

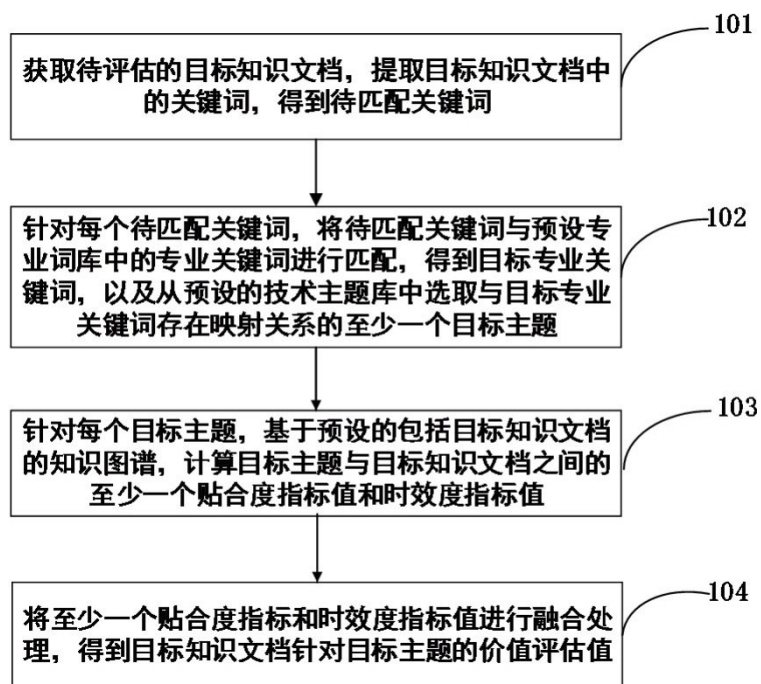


图1

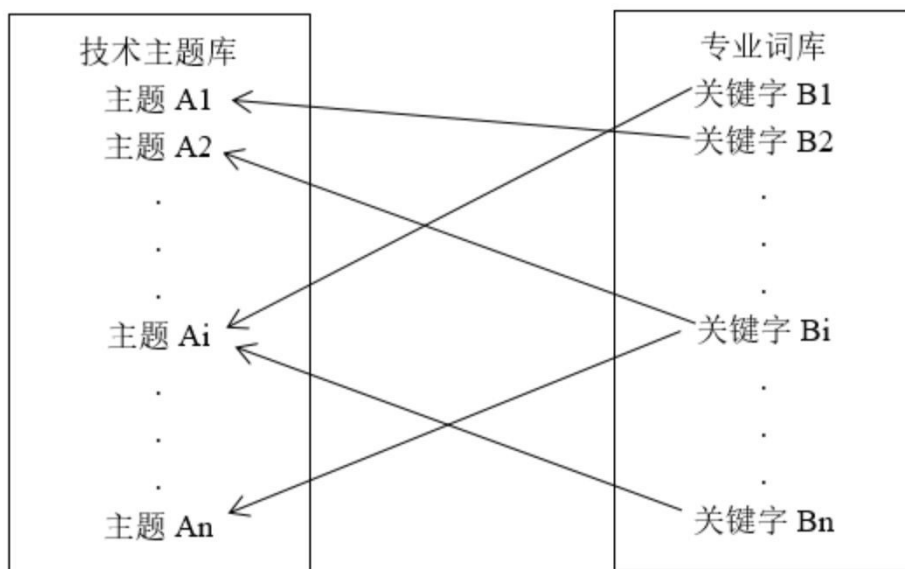


图2

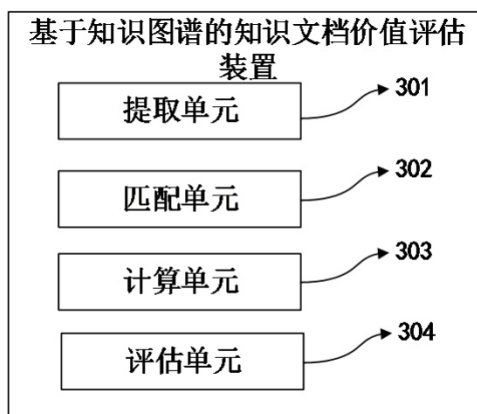


图3

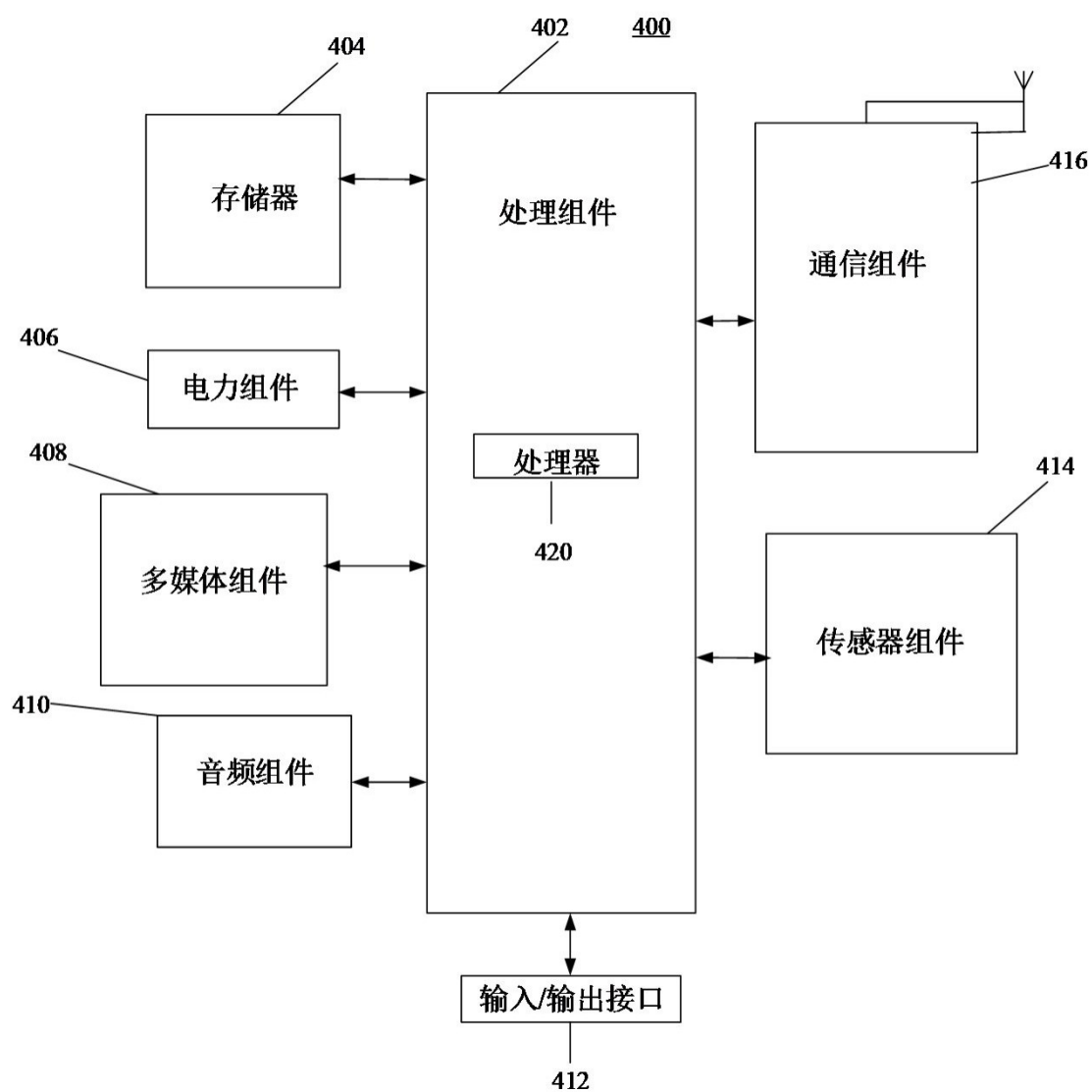


图4